



# HCRI: A Linguistic QA/CA Project

Enrico Trombetta  
Jeff Dalton as supervisor  
Iain Mackie as mentor



# Content

1. Introduction and Problem Statement
2. Limitations of SOTA CAs
3. Research Questions
  1. What is a linguistic question/agent?
  2. What types of multi-term dialogue would users pose?
  3. What data sources can empower such system
  4. How do we answer those questions?
4. PoC
5. Future Work



# Introduction and Problem Statement

Harper Collins Glasgow is interested in improving their products (online mono/bilingual dictionaries) with a QA system to make user interaction less daunting.

The School is interested in providing a domain-specific QA system in the form of a CA that could address linguistic questions.



# QA/CA for dictionaries

# QA/CA for dictionaries

werewolf

English: werewolf

American: werewolf

werewolf

Example sentences

Trends

## Definition of 'werewolf'

# werewolf

Collins COBUILD

Word Frequency ↗

(weəˈwʊlf 🔊)

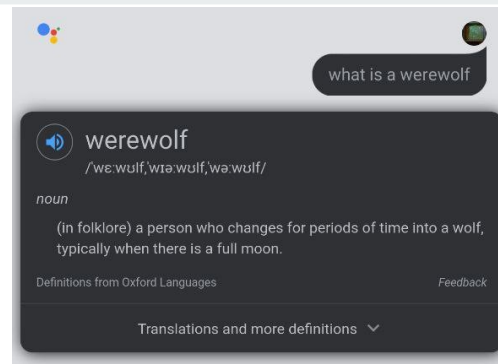
Word forms: plural **werewolves** 🔊

### COUNTABLE NOUN

In stories and films, a **werewolf** is a person who changes into a wolf.

*COBUILD Advanced English Dictionary. Copyright © HarperCollins Publishers*

# QA/CA for dictionaries





werewolf English: werewolf American: werewolf werewolf Example sentences Trends

## Definition of 'werewolf'

# werewolf

Collins COBUILD


(wɛəˈwʊlf )

Word forms: plural **werewolves** 

## COUNTABLE NOUN

In stories and films, a **werewolf** is a person who changes into a wolf.

COBUILD Advanced English Dictionary. Copyright © HarperCollins Publishers

Word Frequency 

# QA/CA for dictionaries

what is a werewolf

**werewolf**  
/wɛːwʊlf, wɪəːwʊlf, wəːwʊlf/  
noun  
(in folklore) a person who changes for periods of time into a wolf, typically when there is a full moon.  
Definitions from Oxford Languages Feedback

Translations and more definitions

what is it's plural

noun, **plural** were-wolves [wair-woolvz, weer-, wur-].  
(in folklore and superstition) a human being who has changed into a wolf, or is capable of assuming the form of a wolf, while retaining human intelligence.  
<https://www.dictionary.com/browse>  
[Werewolf | Definition of Werewolf at Dictionary.com](#)  
About Featured Snippets Feedback

werewolf English: werewolf American: werewolf werewolf Example sentences Trends

## Definition of 'werewolf'

**werewolf**

Collins COBUILD

(weəˈwʊlf )

Word forms: plural **werewolves**

### COUNTABLE NOUN

In stories and films, a **werewolf** is a person who changes into a wolf.

COBUILD Advanced English Dictionary. Copyright © HarperCollins Publishers

Word Frequency

Watch out the  
“it’s”!

# QA/CA for dictionaries

werewolf English: werewolf American: werewolf werewolf Example sentence

## Definition of 'werewolf'

# werewolf

Collins COBUILD

(weəˈwʊlf )

Word forms: plural **werewolves**

### COUNTABLE NOUN

In stories and films, a **werewolf** is a person who changes into a wolf.

COBUILD Advanced English Dictionary. Copyright © HarperCollins Publishers

what is a werewolf

**werewolf**  
/wɛːwʊlf, wɪəːwʊlf, wəːwʊlf/  
noun  
(in folklore) a person who changes for periods of time into a wolf, typically when there is a full moon.  
Definitions from Oxford Languages Feedback

Translations and more definitions

is it a countable noun

<https://www.collinsdictionary.com> ...  
Werewolf definition and meaning | Collins English Dictionary  
werewolf. (weəˈwʊlf ). Word forms: plural werewolves. countable noun. In stories and films, a werewolf is a person who changes into a wolf. COBUILD Advanced English Dictionary.

<https://www.macmillandictionary.com> ...  
WEREWOLF (noun) definition and synonyms | Macmillan Dictionary  
Definition of WEREWOLF (noun): imaginary human changing into wolf during full moon. ... werewolf Definitions and Synonyms. noun countable. UK /weə(r)wʊlf/ ...

what is it's plural

noun, plural were-wolves [wair-woolvz, weer-, wur-].  
(in folklore and superstition) a human being who has changed into a wolf, or is capable of assuming the form of a wolf, while retaining human intelligence.  
<https://www.dictionary.com> · browse  
Werewolf | Definition of Werewolf at Dictionary.com  
About Featured Snippets Feedback

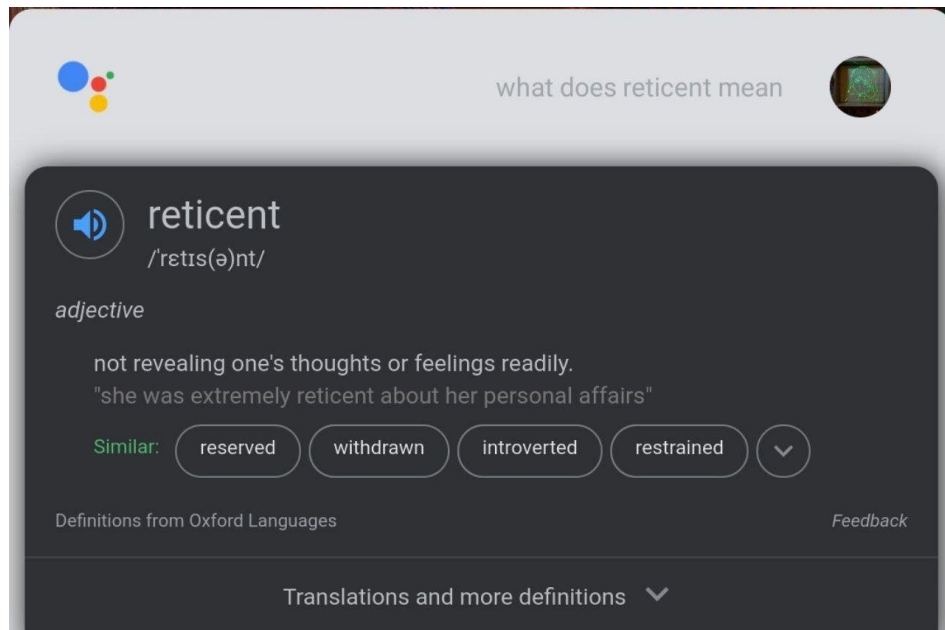
Watch out the  
"it's"!






# Limits of SOTA CAs

# Limits of SOTA CAs



The screenshot shows a Google Assistant interface. At the top, the Google Assistant logo is on the left, and the text "what does reticent mean" is in the center. To the right of the text is a small circular image of a landscape. Below this, a dark grey card displays the word "reticent" in a large font, with a speaker icon to its left. Below the word is the phonetic transcription "/ˈrɛtɪs(ə)nt/". Underneath the transcription, the word "adjective" is written in a smaller font. The definition "not revealing one's thoughts or feelings readily." is shown, followed by an example sentence in quotes: "she was extremely reticent about her personal affairs". Below the definition, the word "Similar:" is followed by four rounded rectangular buttons: "reserved", "withdrawn", "introverted", and "restrained". To the right of these buttons is a small downward-pointing arrow. At the bottom left of the card, it says "Definitions from Oxford Languages", and at the bottom right, it says "Feedback". At the very bottom of the screen, there is a dark grey bar with the text "Translations and more definitions" and a downward-pointing arrow.

what does reticent mean

 **reticent**  
/ˈrɛtɪs(ə)nt/

*adjective*

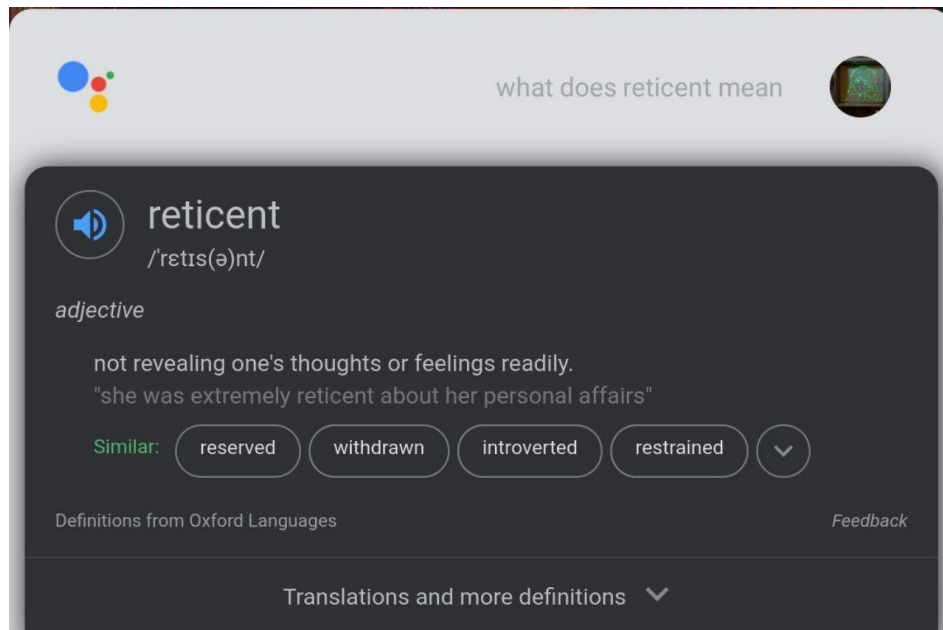
not revealing one's thoughts or feelings readily.  
"she was extremely reticent about her personal affairs"

Similar: reserved withdrawn introverted restrained ▼

Definitions from Oxford Languages [Feedback](#)

Translations and more definitions ▼

# Limits of SOTA CAs



what does reticent mean

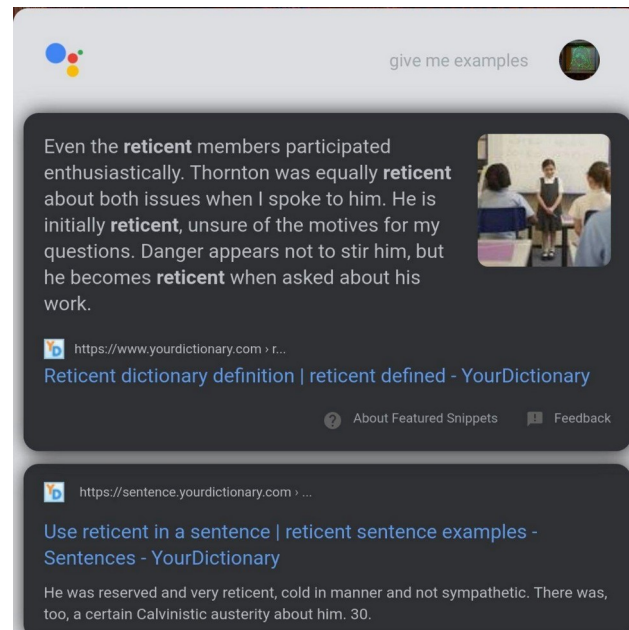
**reticent**  
/'retɪs(ə)nt/  
*adjective*

not revealing one's thoughts or feelings readily.  
"she was extremely reticent about her personal affairs"

Similar: reserved withdrawn introverted restrained

Definitions from Oxford Languages

Translations and more definitions



give me examples

Even the **reticent** members participated enthusiastically. Thornton was equally **reticent** about both issues when I spoke to him. He is initially **reticent**, unsure of the motives for my questions. Danger appears not to stir him, but he becomes **reticent** when asked about his work.

<https://www.yourdictionary.com/r...>  
Reticent dictionary definition | reticent defined - YourDictionary

<https://sentence.yourdictionary.com/...>  
Use reticent in a sentence | reticent sentence examples - Sentences - YourDictionary

He was reserved and very reticent, cold in manner and not sympathetic. There was, too, a certain Calvinistic austerity about him. 30.

# Limits of SOTA CAs

can reticent be a noun

**Reticence** is a **noun** originating from the Latin word **reticent**, meaning “remaining silent.” Someone charged with a crime might be instructed by their lawyer to show **reticence** when being questioned by police.

✓ <https://www.vocabulary.com › retice...>

[reticence - Dictionary Definition : Vocabulary.com](#)

? About Featured Snippets    ! Feedback

give me examples

Even the **reticent** members participated enthusiastically. Thornton was equally **reticent** about both issues when I spoke to him. He is initially **reticent**, unsure of the motives for my questions. Danger appears not to stir him, but he becomes **reticent** when asked about his work.



<https://www.yourdictionary.com › r...>

[Reticent dictionary definition | reticent defined - YourDictionary](#)

? About Featured Snippets    ! Feedback

<https://sentence.yourdictionary.com › ...>

[Use reticent in a sentence | reticent sentence examples - Sentences - YourDictionary](#)

He was reserved and very reticent, cold in manner and not sympathetic. There was, too, a certain Calvinistic austerity about him. 30.

# Limits of SOTA CAs

can reticent be a noun

**Reticence** is a **noun** originating from the Latin word **reticent**, meaning “remaining silent.” Someone charged with a crime might be instructed by their lawyer to show **reticence** when being questioned by police.

✓ <https://www.vocabulary.com › retice...>

[reticence - Dictionary Definition : Vocabulary.com](https://www.vocabulary.com › reticence)

? About Featured Snippets    ? Feedback

show me all possible parts of speech of that word

In any sentence, each word would be one of **NOUN, PRONOUN, VERB, ADVERB, ADJECTIVE, PREPOSITION, CONJUNCTION** and **INTERJECTION**. These 8 grammatical categories are called PARTS OF SPEECH.

8 Parts of Speech	
Adjectives	Adverbs
Conjunctions	Interjections
Nouns	Prepositions
Pronouns	Verbs

Q <https://www.quora.com › What-are-...>

[What are all the possible parts of speech? - Quora](https://www.quora.com › What-are-...)

? About Featured Snippets    ? Feedback

He was reserved and very reticent, cold in manner and not sympathetic. There was, too, a certain Calvinistic austerity about him. 30.

# Limits of SOTA CAs

This is wrong in many ways!

can reticent be a noun

**Reticence** is a **noun** originating from the Latin word **reticent**, meaning “remaining silent.” Someone charged with a crime might be instructed by their lawyer to show **reticence** when being questioned by police.

✓ <https://www.vocabulary.com › retice...>

[reticence - Dictionary Definition : Vocabulary.com](https://www.vocabulary.com › reticence)

? About Featured Snippets    ? Feedback

show me all possible parts of speech of that word

In any sentence, each word would be one of **NOUN, PRONOUN, VERB, ADVERB, ADJECTIVE, PREPOSITION, CONJUNCTION** and **INTERJECTION**. These 8 grammatical categories are called PARTS OF SPEECH.

8 Parts of Speech	
Adjectives	Adverbs
Conjunctions	Interjections
Nouns	Prepositions
Pronouns	Verbs

Q <https://www.quora.com › What-are-...>

[What are all the possible parts of speech? - Quora](https://www.quora.com › What-are-...)

? About Featured Snippets    ? Feedback

He was reserved and very reticent, cold in manner and not sympathetic. There was, too, a certain Calvinistic austerity about him. 30.



# The problem about evaluation

As far as we can tell, nobody has already tried a conversational agent dedicated to linguistics. Furthermore, no known linguistic conversational dataset has ever been compiled, making evaluation more difficult.

Google Home simply forwards the question as a normal Google query. It's able to preserve a bit of context (arguably replaces the 'its' with past keywords) but it breaks quite easily after a few turns.

However, many IR systems work so far quite well because there is plenty of available datasets to query from.

---

# ColLA\*: A Conversational framework for Linguistic Agents

(colla: glue in italian)





# Q1: What is a linguistic question?

From Wikipedia:

**Linguistics** is the scientific study of language.<sup>[1]</sup> It involves the analysis of language form, language meaning, and language in context.<sup>[2]</sup>

If there is plenty of datasets on languages, why is QA on linguistics difficult?

- Linguistics is a ‘meta’ field, as sentences may not follow a traditional structure or diverge radically or merge different grammars or languages in a single utterance!
- Utterances on linguistics are more difficult to find (compared to plain English prose, or even Scientific/Academic English used in research documents)
- We are ‘less’ interested in “encyclopedic definitions”. One might say: “linguistic definitions are the ones that are not inherently encyclopedic”. But what does “encyclopedic” mean here?



## Q1: What is a linguistic question?

(KG-centric definition) A linguistic question is a question that can be answered by a “linguistic” knowledge graph by only leveraging synsets, grammatical properties and syntactic or semantic pointers... where

- Synset: synonym set, i.e. set of meanings associated with “lexemes”. If I replace a word with another one in the same synset the overall meaning does not change (e.g. partner/fiancé/boyfriend)
- Grammatical property: POS, form...
- Semantic pointer: antonym (opposite-of), meronym (is-part-of), hyponym (is-a), hypernym (is-a reversed) ...

Main drawback of this definition: assumes all languages share the same underlying structure. Language philosophers have mixed views.



## Q1: What is a linguistic question?

This allows us to greatly simplify our knowledge graph (no care in embedding other encyclopedic KG properties) at the cost of greatly downscaling the range of questions we can answer to.



## Q2: What types of multi-term dialogue would users pose?

- For example: “define ‘walnuts’”, “how would you say gentrification in German?”, “is luna a male or female noun in Italian?”, “show me synonyms for construction”
- Excluded candidates: purely encyclopedic questions (“what are the other greek gods in the Pantheon?”, “Which countries did Christopher Columbus sail to?”);
  - This goes in contrast with our ‘definition’ in Q1 as it requires more than just linguistic knowledge.
- But this is \*us\* researchers making ‘artificial’ questions, not the end users.



## Q2: What types of multi-term dialogue would users pose?

- Our setting: we took the Google NLQ dataset, preserved the question and id alone and generated a Lucene index to query with Pyserini
  - Used the first 10 'slices' of Google NLQ from the train partition, for a total of 61477 examples
  - stored docvectors + positions
  - Default BM25 parameters:  $k1=0.9, b=0.4$  (we don't really need reranking)
- We hand-picked non-linguistic definitions



## Q2: What types of multi-term dialogue would users pose?

We identified these clusters:

- Definition (“What is the meaning of...?”)
- How is X translated into Y? (“where X is usually a phrase”)
- Questions on grammar rules (“where does the verb go in a german sentence”)

The rest is mainly encyclopedic questions.

Quite disappointing!



## Q2: What types of multi-term

We identified these clusters:

- Definition (“What is the meaning of...?”)

for query: mean

2.69320011138916 what does it mean to be one standard deviation below the mean  
2.6931991577148438 difference between mean arterial pressure and mean systemic pressure  
2.6440000534057617 difference between man search for meaning and man search for ultimate meaning  
2.6439990997314453 what does marx mean by the terms mode of production and means of production  
2.59660005569458 what literary term means a word that has a similar meaning to another word  
2.3547000885009766 the fruit of the spirit and their meaning  
2.35469913482666 the meaning of a raisin in the sun  
2.3546981811523438 this is the story of a girl meaning  
2.3546969890594482 the meaning of the word in the bible  
2.354696035385132 what is the meaning of this symbol #

for query: in urdu

4.719099998474121 what is the meaning of ijma in urdu  
4.719099044799805 what is the meaning of timeline in urdu  
4.574999809265137 what is senate election in pakistan in urdu  
4.57499885559082 what is meant by terrorism in urdu language  
4.574997901916504 what is meant by terrorism in urdu language  
4.5749969482421875 urdu is the official language of which state  
4.574995994567871 essay on importance of quran in urdu language  
4.4394001960754395 one word that describe me meaning in urdu  
4.439399242401123 who can duet with me meaning in urdu  
4.439398288726807 who can duet with me meaning in urdu



## Q2: What types of multi-term dialogue would users pose?

- How does a conversation continue? Does it even continue at all? Are current chatbots suitable for linguistic dialogues?
  - “What does reticent mean” -> “Show me examples” clearly shows some coreference resolution going on, with the need of a context.
  - “What does reticent mean” -> “Show me examples” -> “Can it be used as a noun” -> “What are all the others parts of speech available for that word” breaks apart.
- We do not have an existing “natural” multi-turn dataset on language at hand.






### Q3: What data sources can empower such system

- A good bunch of KGs out there (Wikidata, DBPedia, Freebase, YAGO, BabelNet etc...)
  - but their schemas (or ontologies) diverge radically!
- Unstructured data sources (wiktionary, HC dictionary etc., Wikipedia, other encyclopediae)
  - But they are “hard” to query, especially when looking for very structured information (e.g. morphology)
- We want to achieve a certain level of generality and granularity, able to answer a range of questions in Q2
  - Not trivial. Questions can range from “is ‘produce’ a noun or a verb?” to “give me an example of ‘salir’ in Spanish” to “define Neptune and give me correlated results”.
  - The final KG may have to deal with ambiguities, examples and links to related entities (synonyms, hyperonyms, opposites, word frequency ...). Some level of generality is wished.
  - **Novelty:** forms of adjectives, verbs, nouns etc. as part of the KG. Very useful especially for highly flexed languages





---

## werewolf in a work of fiction (Q30174967)

---

werewolf appearing in a work of fiction

fictional werewolf | fictional lycantrope | werewolf | lycantrope




## werewolf in a work of fiction (Q30174967)

werewolf appearing in a work of fiction

fictional werewolf | fictional lycantrope | werewolf | lycantrope

### Noun

- **S: (n) werewolf**, [wolfman](#), [lycanthrope](#), [loup-garou](#) (a monster able to change appearance from human to wolf and back again)



---

## werewolf in a work of fiction (Q30174967)

---

werewolf appearing in a work of fiction

fictional werewolf | fictional lycantrope | werewolf | lycantrope

### Noun

- **S: (n) werewolf**, [wolfman](#), [lycanthrope](#), [loup-garou](#) (a monster able to change appearance from human to wolf and back again)

### Noun [\[ edit \]](#)

**werewolf** (*plural* [werewolves](#))

1. (*mythology*) A [person](#) who is [transformed](#) or can transform into a [wolf](#) or a wolflike [human](#), often said to transform during a [full moon](#).

### Synonyms [\[ edit \]](#)

- [wolfman](#)
- [lycanthrope](#)
- [man-wolf](#)

### Hyponyms [\[ edit \]](#)

- [wolfwoman](#)

### Derived terms [\[ edit \]](#)

- [werewolfdom](#)



## werewolf in a work of fiction (Q30174967)

werewolf appearing in a work of fiction

fictional werewolf | fictional lycanthrope | werewolf | lycanthrope

### Noun

- **S: (n) werewolf**, [wolfman](#), [lycanthrope](#), [loup-garou](#) (a monster appearance from human to wolf and back again)

dbp:Country

▪ [dbr:England](#)

dbp:grouping

▪ [dbr:Legendary\\_creature](#)

dbp:name

▪ [Werewolf / Lycanthrope](#)

dbp:region

▪ [dbr:Asia](#)  
▪ [dbr:Europe](#)  
▪ [dbr:The\\_Americas](#)  
▪ [dbr:Africa](#)

inform  
orm

dbp:SimilarCreatures

▪ [dbr:Vampire](#)  
▪ [dbr:Therianthropy](#)  
▪ [dbr:Yōkai](#)  
▪ [dbr:Revenant](#)

dbp:SubGrouping

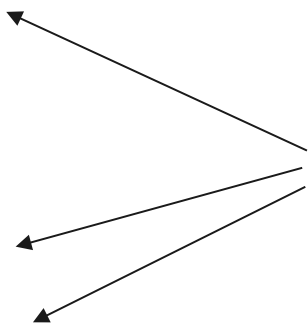
▪ [Lycanthrope](#)

dct:subject

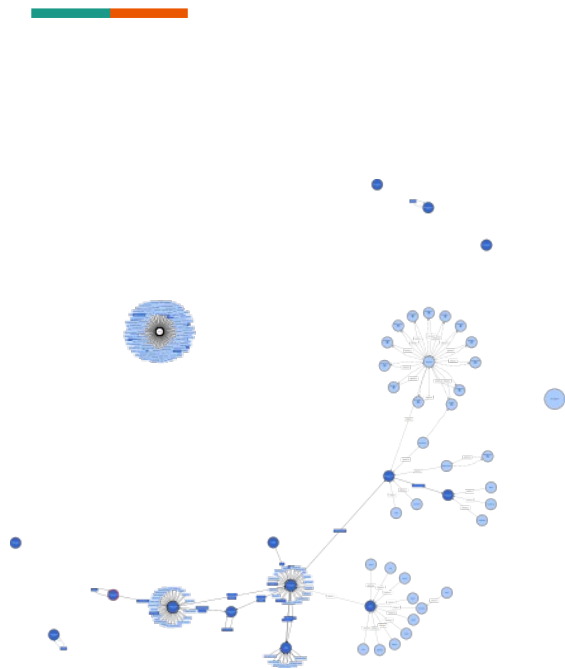
▪ [dbc:Werewolves](#)  
▪ [dbc:Mythic\\_humanoids](#)  
▪ [dbc:Shapeshifting](#)

## Used ontologies in Wikidata Lexemes

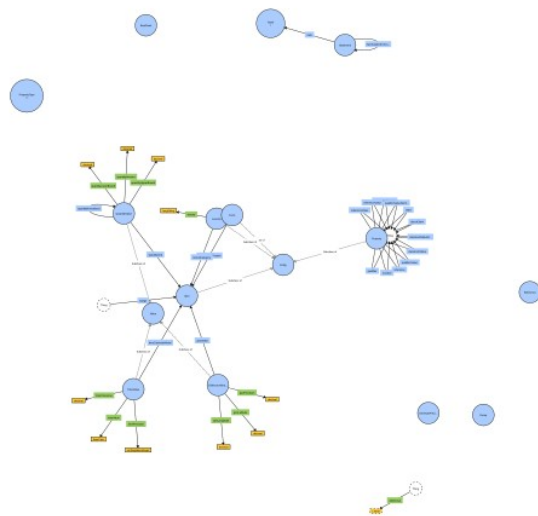
```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix wikibase: <http://wikiba.se/ontology#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix schema: <http://schema.org/> .
@prefix cc: <http://creativecommons.org/ns#> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix data: <https://www.wikidata.org/wiki/Special:EntityData/> .
@prefix s: <http://www.wikidata.org/entity/statement/> .
@prefix ref: <http://www.wikidata.org/reference/> .
@prefix v: <http://www.wikidata.org/value/> .
@prefix wdt: <http://www.wikidata.org/prop/direct/> .
@prefix wdt_n: <http://www.wikidata.org/prop/direct-normalized/> .
@prefix p: <http://www.wikidata.org/prop/> .
@prefix ps: <http://www.wikidata.org/prop/statement/> .
@prefix psv: <http://www.wikidata.org/prop/statement/value/> .
@prefix psn: <http://www.wikidata.org/prop/statement/value-normalized/> .
@prefix pq: <http://www.wikidata.org/prop/qualifier/> .
@prefix pqv: <http://www.wikidata.org/prop/qualifier/value/> .
@prefix pqn: <http://www.wikidata.org/prop/qualifier/value-normalized/> .
@prefix pr: <http://www.wikidata.org/prop/reference/> .
@prefix prv: <http://www.wikidata.org/prop/reference/value/> .
@prefix prn: <http://www.wikidata.org/prop/reference/value-normalized/> .
@prefix wdn: <http://www.wikidata.org/prop/novalue/> .
```



Different ontologies  
imply conceptual  
idiosyncrasies



Lexinfo (“used” in BabelNet)



Wikiba.se for lexemes, used in Wikidata





# BabelNet

## BabelNet

---

From Wikipedia, the free encyclopedia

**BabelNet** is a [multilingual](#) lexicalized [semantic network](#) and [ontology](#) developed at the [NLP](#) group of the [Sapienza University of Rome](#).<sup>[1][2]</sup> BabelNet was automatically created by linking Wikipedia to the most popular computational [lexicon](#) of the [English language](#), [WordNet](#). The integration is done using an automatic mapping and by filling in lexical gaps in resource-poor [languages](#) by using [statistical machine translation](#). The result is an [encyclopedic dictionary](#) that provides [concepts](#) and [named entities](#) lexicalized in many languages and connected with large amounts of [semantic relations](#). Additional lexicalizations and definitions are added by linking to free-license wordnets, OmegaWiki, the English [Wiktionary](#), [Wikidata](#), [FrameNet](#), [VerbNet](#) and others. Similarly to WordNet, BabelNet groups [words](#) in different languages into sets of [synonyms](#), called *Babel synsets*. For each Babel synset, BabelNet provides short definitions (called [glosses](#)) in many languages harvested from both WordNet and Wikipedia.

# BabelNet

## BabelNet

From Wikipedia, the free encyclopedia

**BabelNet** is a [multilingual](#) lexicalized [semantic network](#) developed at the [NLP](#) group of the [Sapienza University of Rome](#). BabelNet was automatically created by linking popular computational [lexicon](#) of the [English language](#) with integration is done using an automatic mapping of gaps in resource-poor [languages](#) by using [statistical methods](#). The result is an [encyclopedia](#) that contains [named entities](#) lexicalized in many language amounts of [semantic relations](#). Additional lexicons are added by linking to free-license wordnets such as [Wiktionary](#), [Wikidata](#), [FrameNet](#), [VerbNet](#) and [WordNet](#), BabelNet groups [words](#) in different [synonym sets](#), called *Babel synsets*. For each Babel synset, it provides short definitions (called [glosses](#)) in multiple languages from both WordNet and Wikipedia.

English Arabic Chinese French German Greek Hebrew Hindi Italian

bn:00052113n • NOUN • Concept • Categories: Monsters, Mythic humanoids, Mythological canines, Shapeshifting...

EN werewolf • lycanthrope • wolfman • loup-garou

A monster able to change appearance from human to wolf and back again WordNet

More definitions

IS A	mythical creature • transmogrification • fabulous creature
HAS KIND	Werewolf • vrykolakas • Pricolici
HAS INSTANCE	Oz • Wolfsbane • The Captain
DESCRIBED BY SOURCE	Brockhaus and Efron Encyclopedic Dictionary
HAS QUALITY	lycanthropy
MEDICAL CONDITION	lycanthropy

# BabelNet

## BabelNet

From Wikipedia, the free encyclopedia

**BabelNet** is a **multilingual** lexicalized **semantic network** developed at the **NLP** group of the **Sapienza University of Rome**. BabelNet was automatically created by linking together popular computational **lexicon** of the **English language**. The integration is done using an automatic mapping between concepts in different languages by using statistical methods. The result is an **encyclopedia** that contains **named entities** lexicalized in many languages. BabelNet contains large amounts of **semantic relations**. Additional lexical information is added by linking to free-license wordnets such as **Wiktionary**, **Wikidata**, **FrameNet**, **VerbNet** and **WordNet**. BabelNet groups **words** in different languages that share the same meaning, called **Babel synsets**. For each Babel synset, BabelNet provides short definitions (called **glosses**) in different languages from both WordNet and Wikipedia.

EnglishArabicChinese

bn:00052113n • NOUN • Concept •

EN

werewolf

• lyc

A monster able to change appearance and shape.

+

More definitions

IS A

mythical creature

HAS KIND

Werewolf

HAS INSTANCE

Oz • Wolf

DESCRIBED BY SOURCE

Brockhaus

HAS QUALITY

lycanthropic

MEDICAL CONDITION

lycanthropy

Sources

WordNet senses

EN werewolf<sup>1</sup>, wolfman<sup>1</sup>, lycanthrope<sup>1</sup>, loup-garou<sup>1</sup>

Wikipedia page titles

EN werewolf

Wikipedia redirections

EN Ligahoo, Loup-garou, Loup Garou, Loup garou, Lycanthrope, Lycanthropes, Lycanthropia, Lycanthropic, Lycanthropism, Lycanthrophy, Lykanthropos, Lykánthropos, Lyncanthrope, Nick Demase, Pou-garou, Ruvaush, The Book of Werewolves, User:F0xx0rzz, Valkolak, Varcolac, Varkolak, Voirloups, Vukodlak, Vulkodlak, Vårkolak, Vålkolak, Warewolf, Wearwolf, Wearwolfs, Wearwolves, Wer-wolf, Were-lions, Were-wolf, Were-wolves, Were wolf, Were wolfs, Werewolfs, Werewolve, WereWolves, Werewolves, Wherewolf, Wolf-man, Wolf-Man, Wolf man, Wolfman, Wolfmen, Λυκάνθρωπος

Wiktionary senses

EN lycanthrope, werewolf, wolfman

Wikidata items

EN lycanthrope, werewolf

# BabelNet

## BabelNet

From Wikipedia, the free encyclopedia

**BabelNet** is a **multilingual** lexicalized **semantic** network developed at the **NLP** group of the **Sapienza University of Rome**. BabelNet was automatically created by linking popular computational **lexicon** of the **English language**. The integration is done using an automatic mapping between the **English language** and **resource-poor languages** by using **statistical machine translation**. The result is an **encyclopedia** that **integrates** **named entities** **lexicalized** in many **languages** and **amounts of semantic relations**. Additional **lexical resources** are added by linking to free-license wordnets like **Wiktionary**, **Wikidata**, **FrameNet**, **VerbNet** and **WordNet**. BabelNet groups **words** in different **synonyms**, called **Babel synsets**. For each **Babel synset**, it provides short definitions (called **glosses**) in many **languages** from both **WordNet** and **Wikipedia**.

English
Arabic
Chinese

bn:00052113n • NOUN • Concept •

EN werewolf

A monster able to change appearance

More definitions

IS A mythical creature

HAS KIND Werewolf

HAS INSTANCE Oz • Wolf

DESCRIBED BY SOURCE Brockhaus

HAS QUALITY lycanthropy

MEDICAL CONDITION lycanthropy

transformation

1.5K

fabulous creature

vrykolakas

1.2K

wolf

wulfsbane (Lupinus)

Louis Adore Kahn

werewolf

Wikipedia redirections

Ligahoo, Loup-garou, Loup Garou, Loup garou, Lycanthrope, Lycanthropes, Lycanthropia, Lycanthropic, Lycanthropism, Lycanthrophy, Lykanthropos, Lykánthropos, Lyncanthrope, Nick Demase, Pou-garou, Ruvaush, The Book of Werewolves, User:F0xx0rzz, Valkolak, Varcolac, Varkolak, Voirloups, Vukodlak, Vulkodlak, Vårkolak, Vålkolak, Warewolf, Wearwolf, Wearwolfs, Wearwolves, Wer-wolf, Were-lions, Were-wolf, Were-wolves, Were wolf, Were wolfs, Werewolfs, Werewolve, WereWolves, Werewolves, Wherewolf, Wolf-man, Wolf-Man, Wolf man, Wolfman, Wolfmen, Λυκάνθρωπος

Wiktionary senses

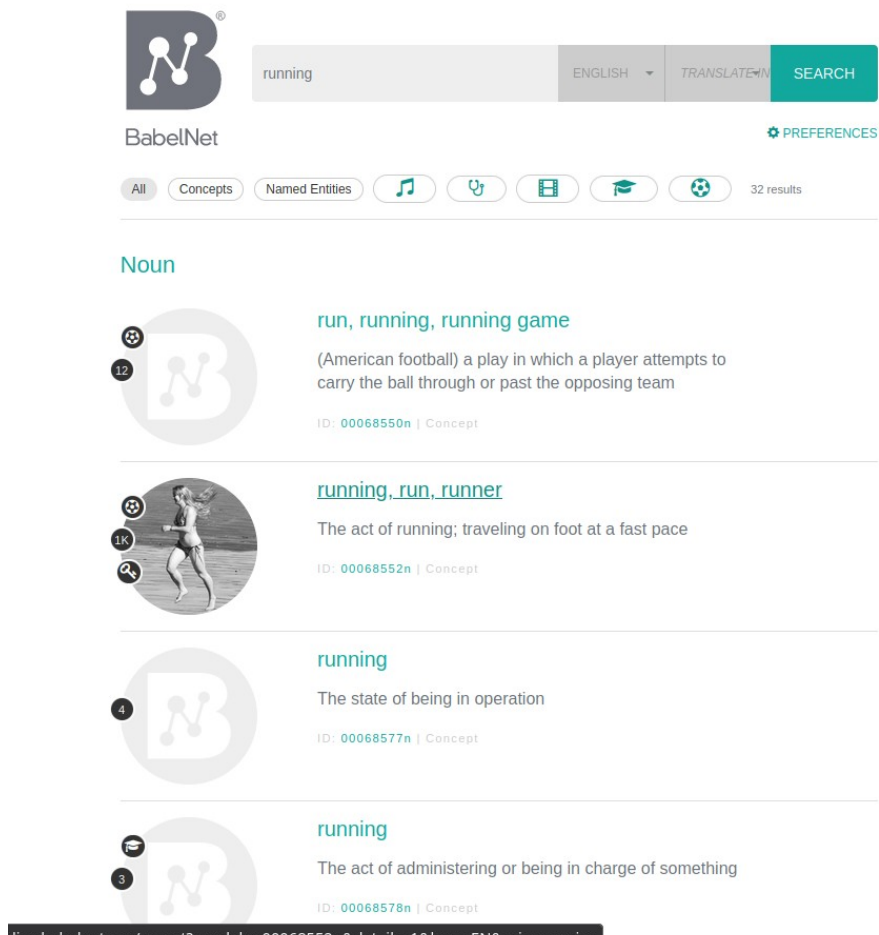
lycanthrope, werewolf, wolfman

Wikidata items





lycanthrope, werewolf

# Forms are a no-go

Form-based search gives bad results on BabelNet, compared to wiktionary.



The screenshot shows the BabelNet search interface. At the top, there is a search bar with the word "running" entered. To the right of the search bar are buttons for "ENGLISH", "TRANSLATE IN", and "SEARCH". Below the search bar, the BabelNet logo is visible, followed by a "PREFERENCES" link. A row of filters is shown: "All", "Concepts", "Named Entities", and several icons representing different media types (music, video, image, etc.). The "All" filter is selected, and the text "32 results" is displayed. The results are categorized under the heading "Noun". There are four results listed, each with a BabelNet icon, a count, a title, a description, and an ID.

Count	Image	Concept	Description	ID
12		run, running, running game	(American football) a play in which a player attempts to carry the ball through or past the opposing team	00068550n   Concept
1K		running, run, runner	The act of running; traveling on foot at a fast pace	00068552n   Concept
4		running	The state of being in operation	00068577n   Concept
3		running	The act of administering or being in charge of something	00068578n   Concept

# Forms are a no-go

Form-based search gives bad results on BabelNet, compared to wiktionary.



BabelNet

No result found for **suonavo** in BabelNet.

The screenshot shows the BabelNet search interface. At the top is the BabelNet logo and a search bar containing the word 'running'. To the right of the search bar are buttons for 'ENGLISH', 'TRANSLATE IN', and 'SEARCH'. Below the search bar is a 'PREFERENCES' link. A row of filters includes 'All', 'Concepts', 'Named Entities', and several icons representing different types of entities. The search results are displayed under the heading 'Noun'. The first result is for 'run, running, running game', with a description: '(American football) a play in which a player attempts to carry the ball through or past the opposing team'. Below this is the ID '00068550n' and the label 'Concept'. The second result is for 'running, run, runner', with a description: 'The act of running: traveling on foot at a fast pace'. Below this is the label 'Italian' and an 'edit' link. The third result is for 'Verb', with an 'edit' link. Below this is the word 'suonavo'. The fourth result is for '1. first-person singular imperfect indicative of suonare', with the word 'running' below it. Below this is the description 'The act of administering or being in charge of something' and the ID '00068578n' and the label 'Concept'.

running

ENGLISH TRANSLATE IN SEARCH

BabelNet

PREFERENCES

All Concepts Named Entities

Noun

run, running, running game

(American football) a play in which a player attempts to carry the ball through or past the opposing team

ID: 00068550n | Concept

running, run, runner

The act of running: traveling on foot at a fast pace

Italian [edit]

Verb [edit]

suonavo

1. first-person singular imperfect indicative of suonare

running

The act of administering or being in charge of something

ID: 00068578n | Concept



# Our final choices for KG data sources

## BabelNet

- Embeds WordNet, VerbNet, YAGO, even Wiktionary sometimes - but also links with corresponding encyclopedic pages (e.g. dbpedia, wikipedia, wiktionary etc.)
- Very linguistic knowledge graph
- Also supports *some* translation
- Closed source, Java-only API, not much introspectable or documented Lucene index.
- Limited to nouns, verbs, adjectives and adverbs

## Wiktionary

- Has quite more entries than BabelNet
- Virtually every English word type is available, including prefixes)
- Users may still want word senses and actual language usages and examples
- Includes quotations and form templates
- More control on the parsing
- Wiktextextract is more python-friendly
- The parser is not perfect (but neither BabelNet data is)



# Knowledge Graph building

- We leverage BabelNet and Wiktionary as much as we can,
  - Sometimes we need to perform entity linking between senses and subsenses
  - ... or choose which POS taxonomy to use (beware, political decision!)
- We sometimes use other data sources:
  - Wikidata for a list of form types - basically anything useful BabelNet scrapped
- The focus is not completeness, but QA! We pruned what we deemed unnecessary.





## Our architecture (broad overview)

### Wiktionary extraction

Very full, complete dictionary with > 1M entries for the English language.  
We leverage a patched version of Tatu Ylonen's wiktextract

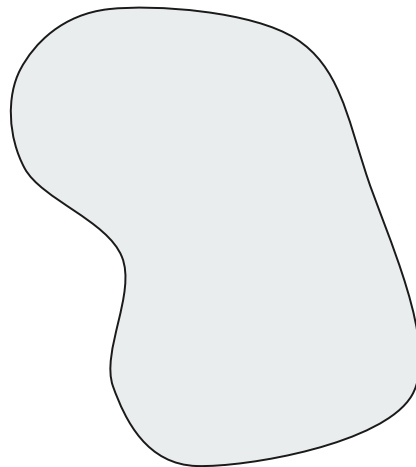
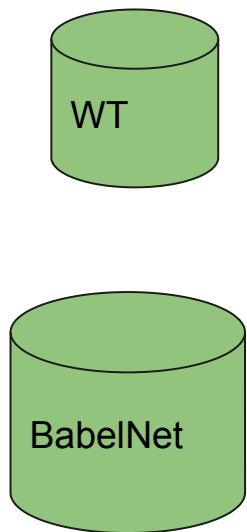
### Entity Linking/Entity Disambiguation with BabelNet

Associate senses with synsets. Synsets present more chances for sense clustering

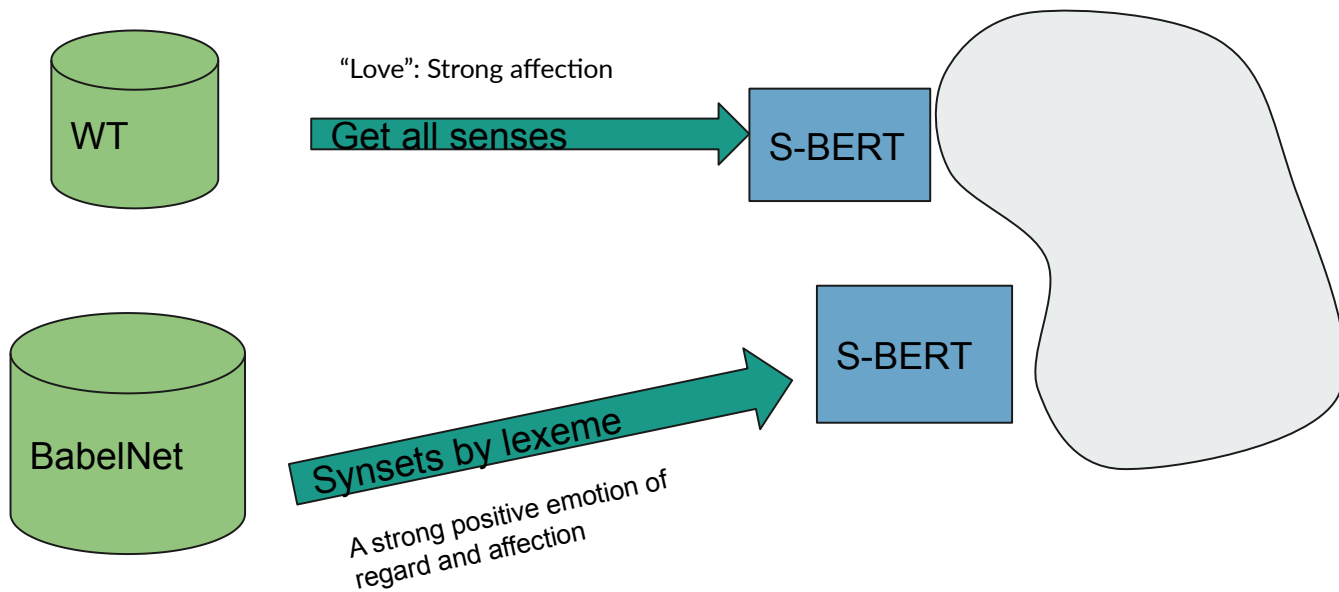
### RDF generation/serialization

Result can be thus imported into standard RDF graph visualizers and processors, e.g. SPARQL engines; or into Lucene documents for typical IR.

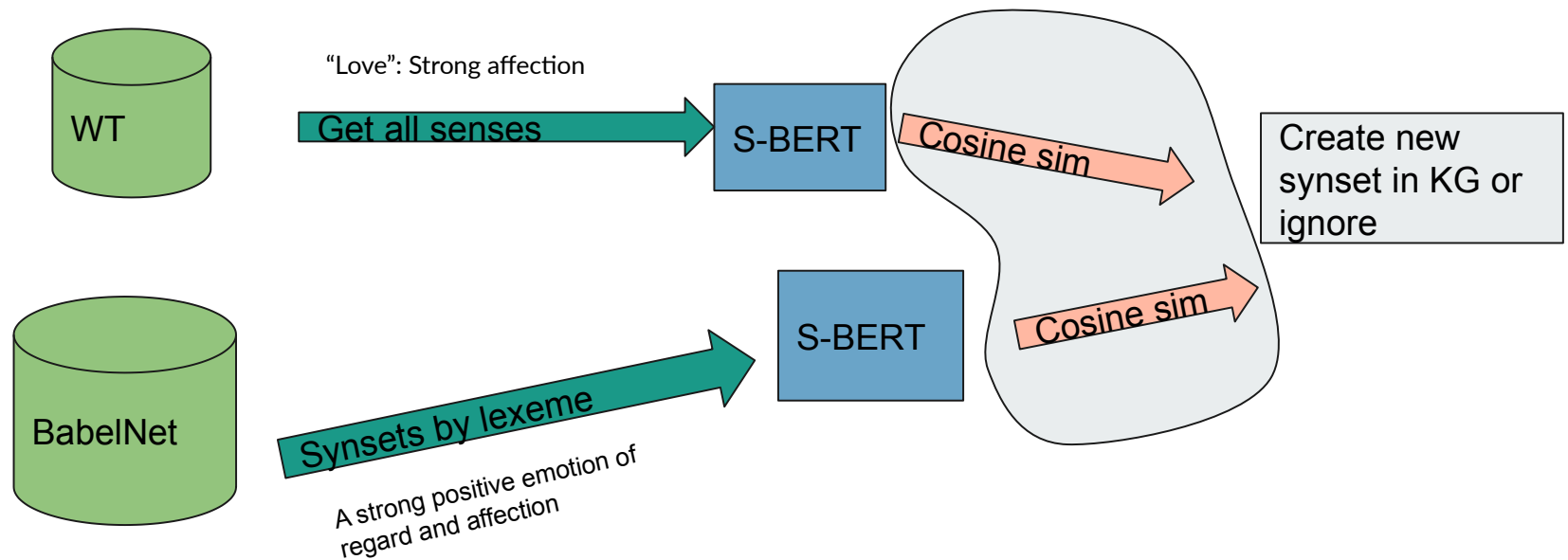
# BabelNet entity linking



# BabelNet entity linking



# BabelNet entity linking





# Knowledge Graph Endpoints

How do we query the Knowledge Graph?

- The graph has been serialized into RDF triples (e.g. Turtle) which can then be imported into any RDF-powered application (e.g. Python's RDFlib) or SPARQL engines
  - <http://knowledge-glue-fuseki-jeffstudentsproject.ida.dcs.gla.ac.uk/>
- Lucene-based search. Good for fuzzy search:
  - Searching for “rose” could match the main lexeme but also slightly related words like “rosemary” - or even a gloss of something else (“bunch”, as in “bunch of roses”)



## Q4: how do we answer linguistic questions?

- Existing models like BART, XL etc. are trained on casual or encyclopedic topics, not linguistic ones!
- Actually, we lack a dataset to start with, let alone training/finetuning.
- We adopt an intent/slot based question translation on a manually annotated small dataset of 50 utterances.
  - Definitely not the “best” way to go, but offers a decent baseline for future work.
- We provide a baseline for an answer, not really an answer.
- Simple idea: we translate natural questions to intents, then we translate those intents into SPARQL queries.



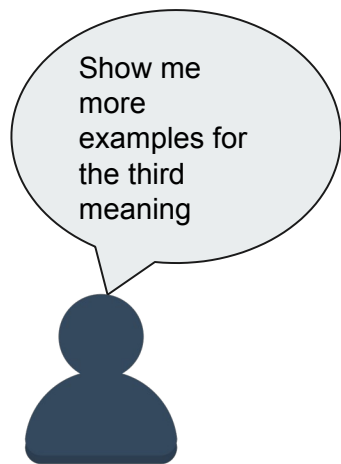
# Intent translation

Our baseline is a context-free grammar with > 140 rules and supports 4 intents:

- Definition intent: fetch definitions for a given noun phrase.
  - “define walnuts”, “what is a palombar”, “tell me what an airplane is”
- Filter intent: give a “projection” of a previous definition entity.
  - “What are possible examples?”, “what is the third-person present form of live?”, “show me more senses”
- Related intent: give related word meanings.
  - “What are possible synonyms?”, “show me some derived words”
- Decide intent: like filter, but answers binary question (yes/no).
  - “Can reticent be used as a noun”?



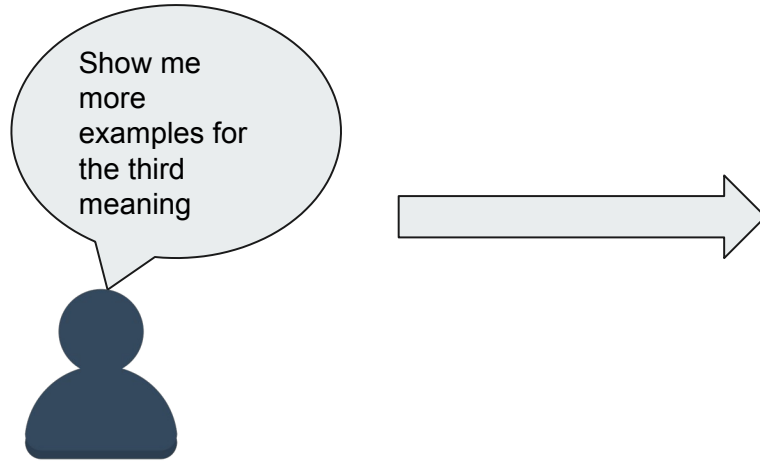
# Intent translation example





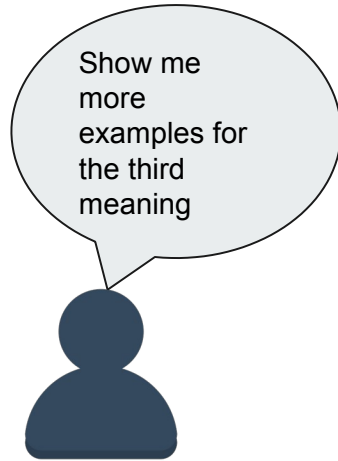


## Intent translation example





# Intent translation example

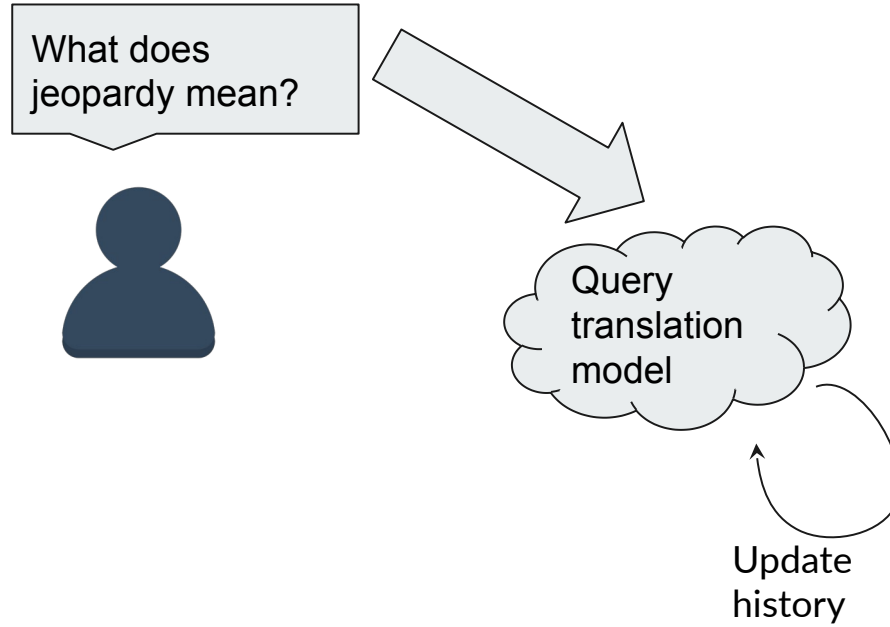


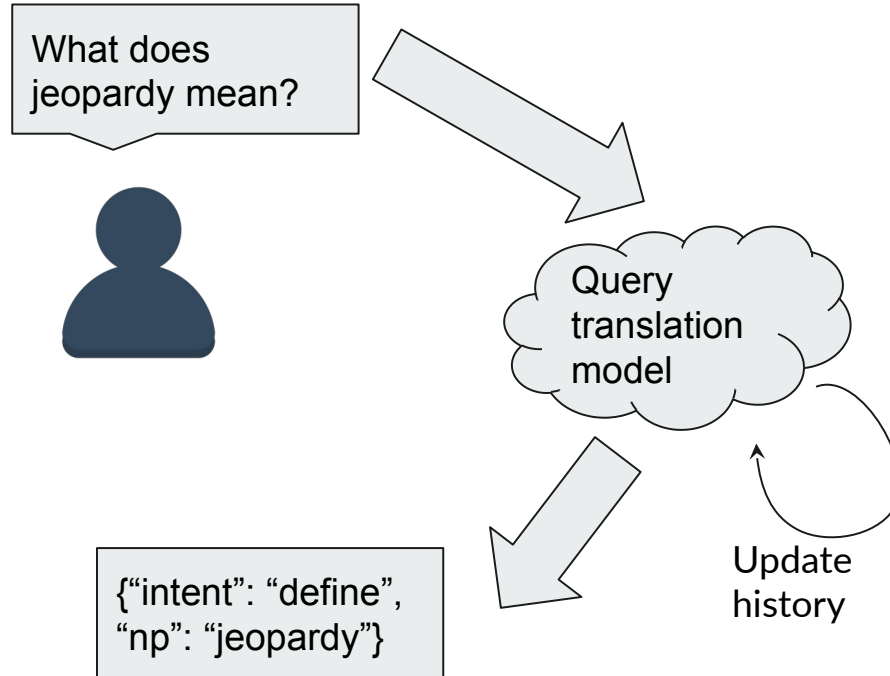
```
{  
  "intent": "filter",  
  "filtertype": "number",  
  "value": 1,  
  "variant": "examples"  
}
```

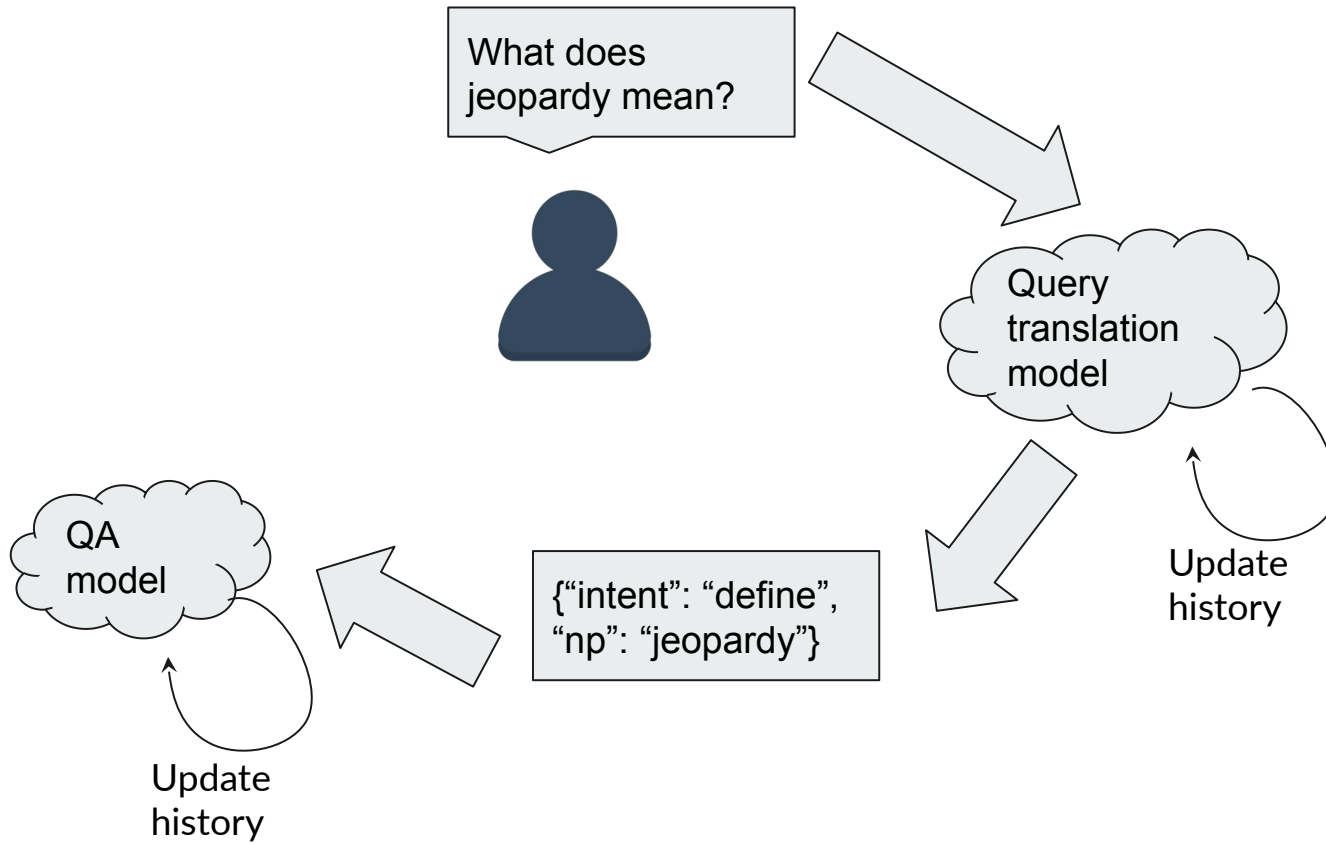


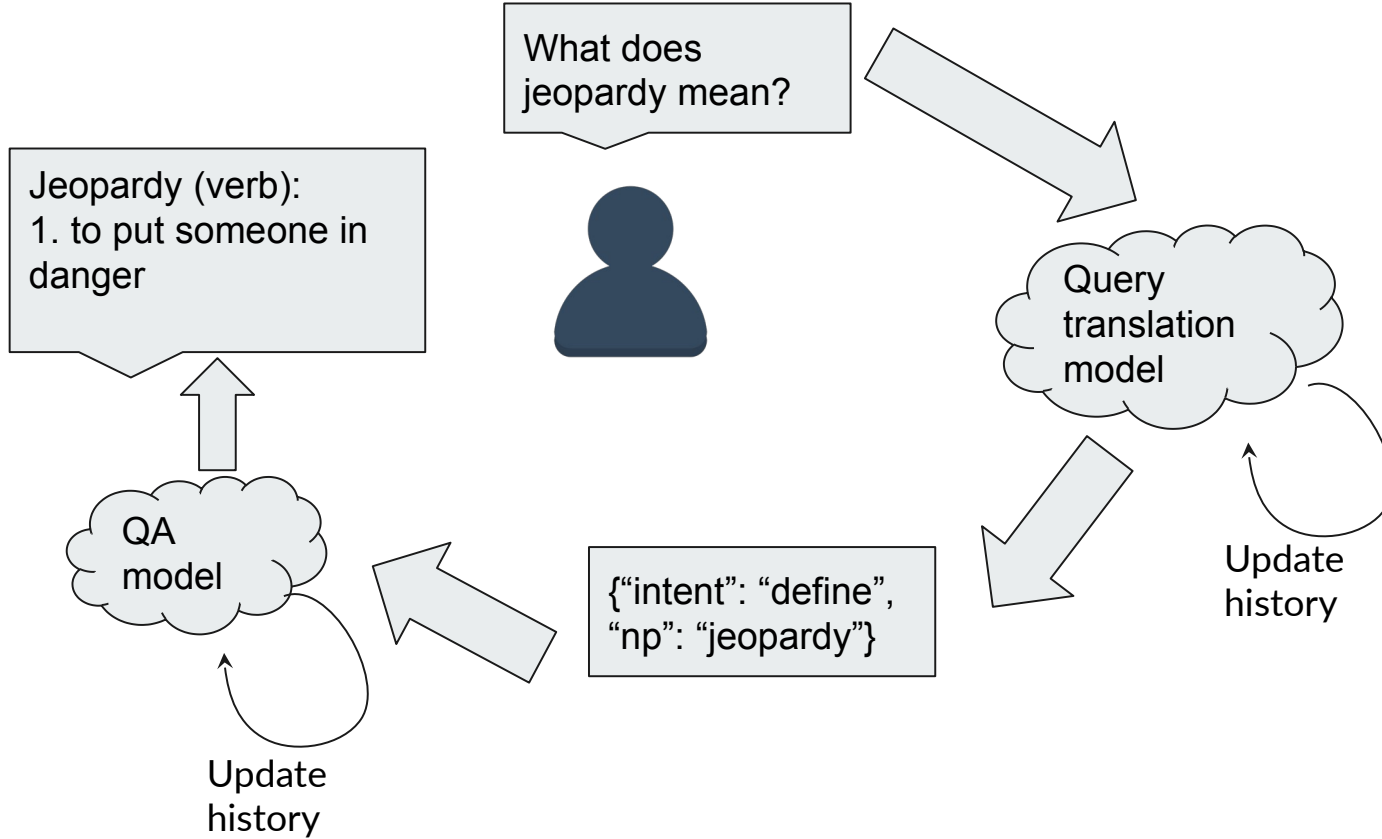
What does  
jeopardy mean?









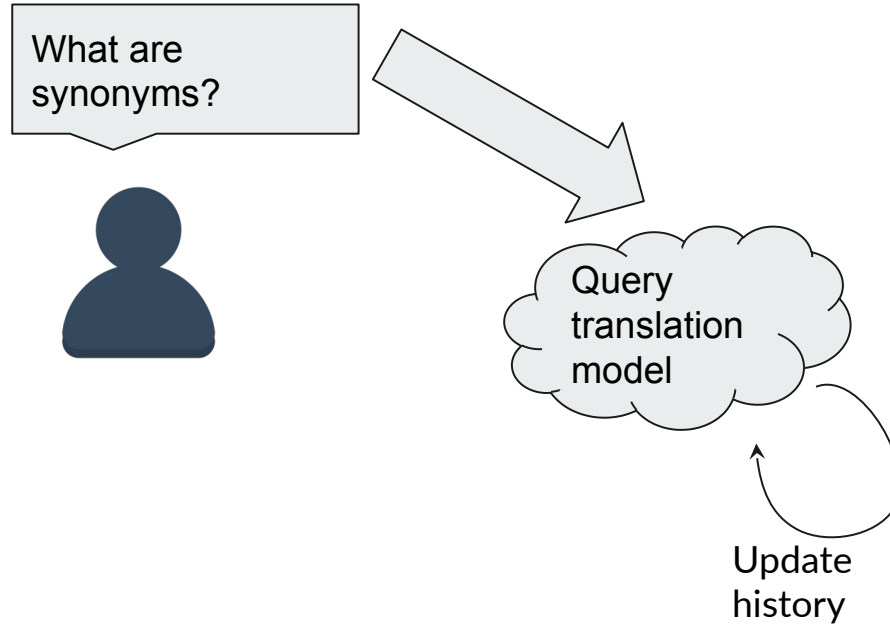


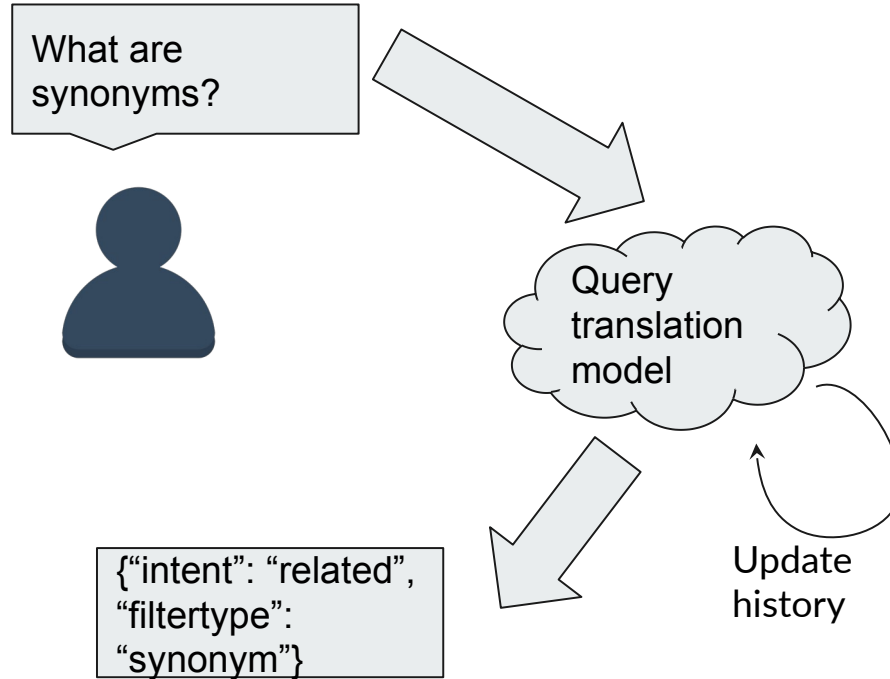


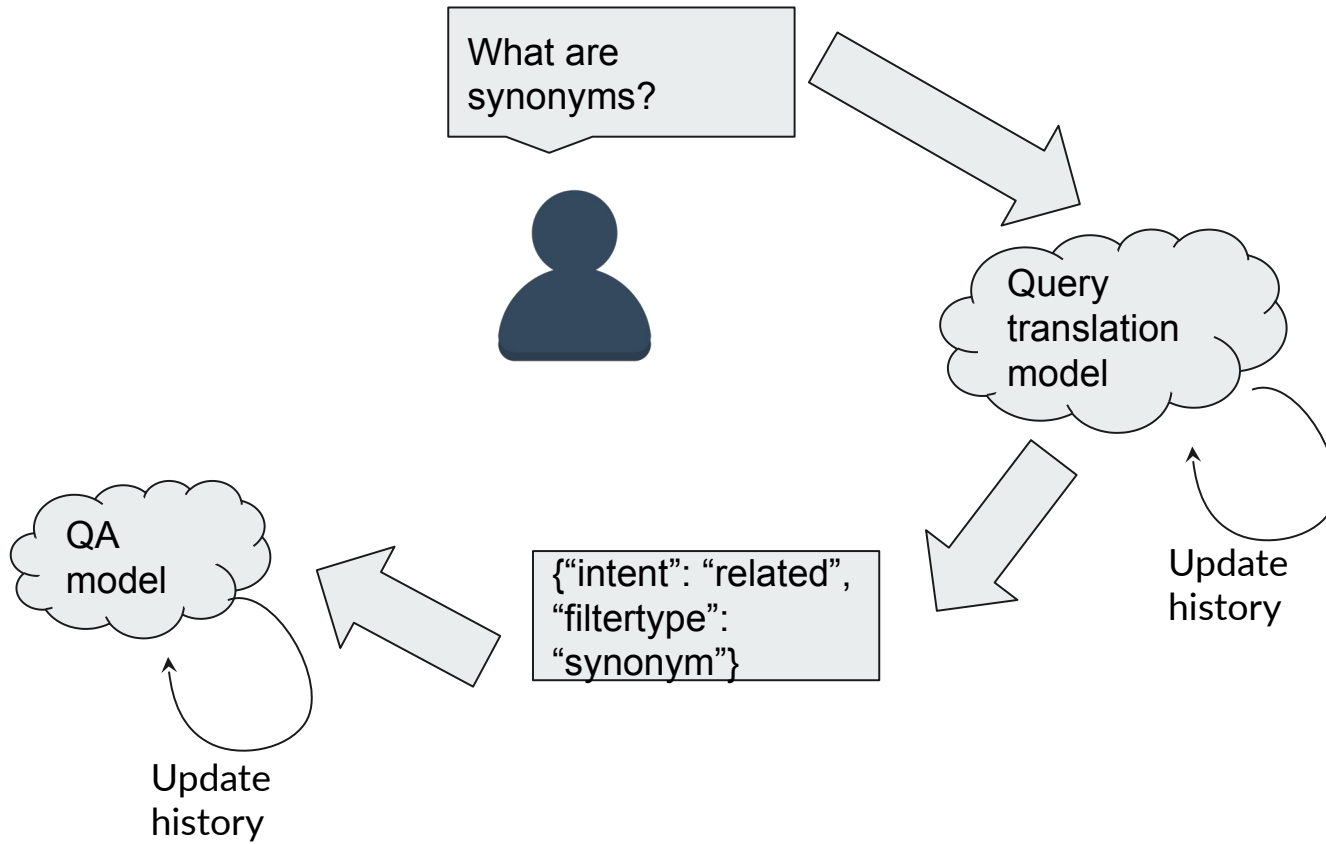


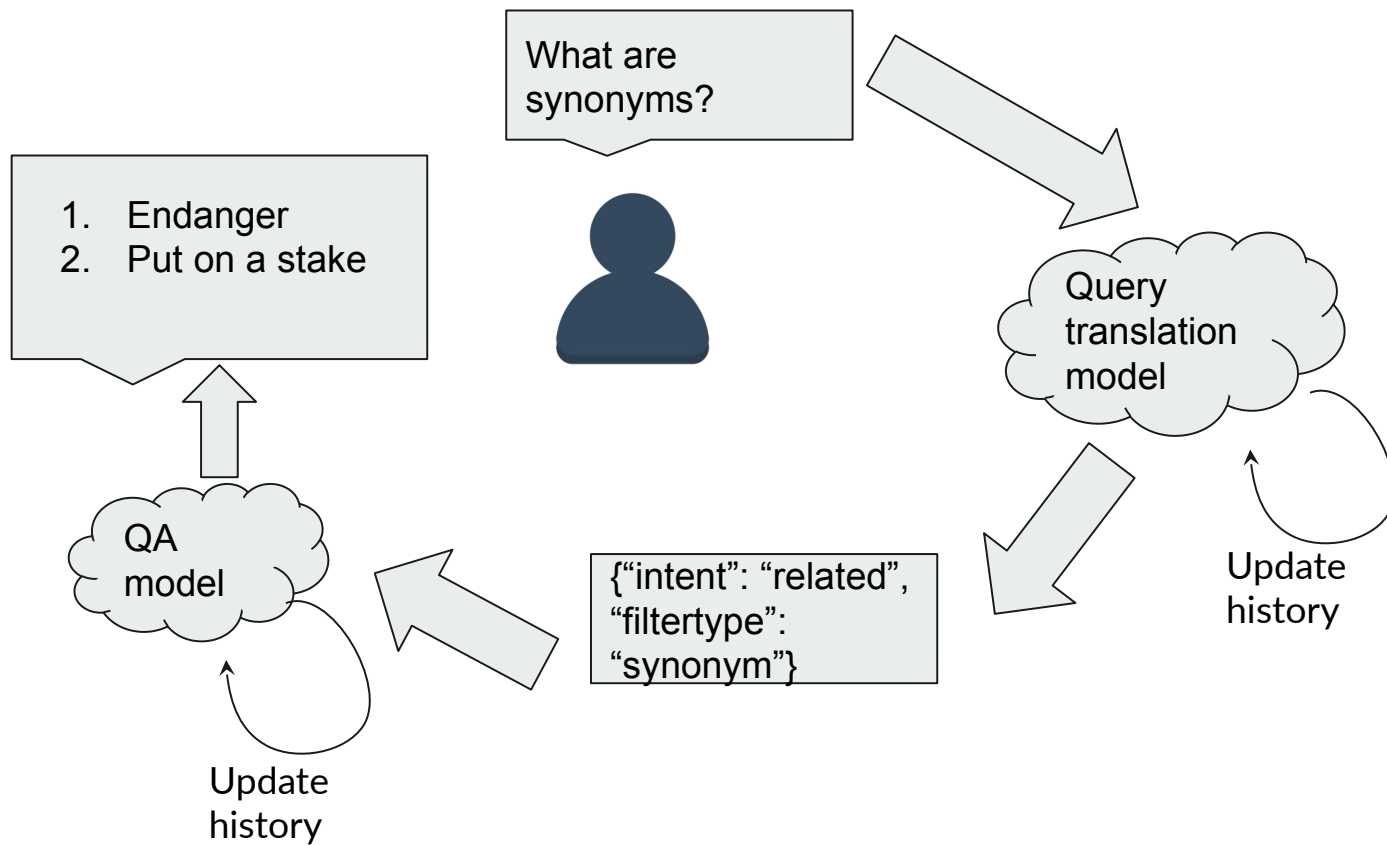
What are  
synonyms?













# PoC

<http://knowledge-glue-webui-jeffstudentsproject.ida.dcs.gla.ac.uk/>

- Chatbot
- Fuzzy searcher
- KG visualizer
  - You can also download an IRI file of the adopted KG ontology and visualize it with <http://www.visualdataweb.de/webvowl> )



## Future work

- Extract etymologies, phonemes
- Support multiple languages for word or 'small phrase' translation. The given data sources are rich enough to afford this, but form structure diverges by a lot.
- Use seq2seq models for NLQ  $\leftrightarrow$  intent translation, or even for direct KG query translation.
- Use better entity linking mechanisms? S-BERT with LSH? Or a different model altogether?



**So long, and thanks for all the fish**

---